NBSIR 81-2222

Some Conservative Statistical Approaches for Presenting Interlaboratory N.D.A. Enrichment Measurements

Clifford H. Spiegelman

Center for Applied Mathematics
National Engineering Laboratory
U.S. Department of Commerce
National Bureau of Standards
Washington, DC 20234

March 1981

Final Report

Prpared for
Nuclear Regulatory Commission
ington, DC 20555

NBSIR 81-2222

SOME CONSERVATIVE STATISTICAL APPROACHES FOR PRESENTING INTERLABORATORY N.D.A. ENRICHMENT MEASUREMENTS

Clifford H. Spiegelman

Center for Applied Mathematics
National Engineering Laboratory
U.S. Department of Commerce
National Bureau of Standards
Washington, DC 20234

March 1981

Final Report

Prpared for
U.S. Nuclear Regulatory Commission
Washington, DC 20555

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

Some Conservative Statistical Approaches for Presenting Interlaboratory N.D.A. Enrichment Measurements

By Clifford Spiegelman

SECTION 1. Introduction.

Realistic evaluation of interlaboratory comparisons of measurements on prototype and primary standards is important for establishing uniformity of reported results. The emphasis in this paper is on multichannel analyzer count measurements, such as those obtained from gamma spectroscopy. These measurements and other nondestructive assay (N.D.A.) measurements are used to measure the enrichment (percentage) of special materials, such as U_3O_8, in the standards. The measurements are based on an estimate of the total number of counts from the special material. This estimate is called the peak area calculation. There is no widespread agreement on the correct formulas to compute peak area, see [1]. Thus when different laboratories measure prototype standards (and hence possibly different formulas are used for computing peak area) it is difficult to distinguish differences among the standards from superficial differences caused by variations in the peak area formulas used by different laboratories, see [1].

The central issue to be considered is whether or not M proposed standards have enrichments sufficiently similar to be certified with the same enrichment level. Several statistical tests of similarity are considered. For two of these tests, algorithms are provided for finding linear peak area formulas which make the standards appear either most or least homogeneous. These two tests are the usual F-statistic and the standardized range. This report provides specific procedures for finding the linear peak area formulas that are best and worst in the sense that they make the prototype standards look most or least homogeneous.

The measurements are a vector of counts. Each entry in the vector is called a channel. A channel contains an integer count of the number of particles at nearly a fixed energy level striking the counter. Elevated counts in certain channels indicate the presence of particular radioisotopes. However, due to residue from past analysis and imperfect counter resolution, adjustments for "background" must be made to the raw counts. Adjusted counts are used to provide more accurate estimates of the amounts of radioisotopes present. One of these adjustments is the peak area computation.

Two components contribute to the peak area computation. One component is counts from the target radioisotope, the other is background. The peak area calculation attempts to find the shaded area in Figure 1. This area is estimated by subtracting an estimate of the area due to background (shown by dots) from the estimate of total area in the peak region. The area due to background is estimated from counts in adjacent "background region" channels. The estimates of area we consider here are linear and are described in detail later in this section. These are the most common type of area estimates, see [1]. Some notation is given below.

Let I_j denote the number of counts in channel j from the target radioisotope.

Let β_j denote the number of background counts in channel j. Let J_1 denote the set of indices of the channels in the peak region and J_2 denote the set of indices of the channels in the purely background region. Without loss of generality take $J_1 = \{1, \ldots, j_1\}$ and $J_2 = \{j_1 + 1, \ldots, j_2\}$. (Assume these regions are given and fixed throughout the remainder of this paper.) Typically J_1 and J_2 are not overlapping. We assume that there are M proposed standards measured by each laboratory. Since different multichannel analyzers

may have different characteristics such as different geometrical construction it does not make sense to seek a fixed peak area formula for all laboratories. Instead, different laboratories' measurements will be dealt with individually. All the notation that follows is for a single laboratory's measurements.

The observables are:

$$Y_j = \begin{cases} I_j + \beta_j & j \in J_1 \\ \beta_j & j \in J_2 \end{cases}$$

The peak area (PA) calculations we consider are linear. Specifically:

$$PA = \sum_{j \in J_1} w_j Y_j + \sum_{j \in J_2} w_j Y_j$$

Since the weights $w_j = 0$ for $j = 1, 2, \ldots, j_2$ would produce a zero peak area for all standards a constraint must be placed on the weights. Otherwise one may get the trivial vector $W^t = (w_1, \ldots, w_{j_2}) = (0, 0, \ldots, 0)$ for the weights that make the standards appear most homogeneous. The constraint used here is

$$\sum_{1}^{j_1} w_j = 1$$

which is standard in numerical quadrature. Since the statistics used in the next section are scale invariant the value, 1, serves only to rule out the vector $W = \underset{\sim}{0}$. In addition it should be noted that nearly all peak area rules can be written as $[W(y)]^t y$ where the vector $W(y)$ is some (often nonlinear) function of $(y_1, \ldots, y_{j_2}) = y^t$, see [1].

Replicate measurements are usually made for each standard. The resulting peak areas are then compared by an analysis of variance procedure, see [2]. Let

μ_k denote the mean peak area for the k-th standard. Let $\bar{\mu}$ denote the average value of μ_k over all the M standards. One possible measure of homogeneity for the M standards is $\Delta^2 = \sum_{1}^{M} (\mu_k - \bar{\mu})^2$. If Δ is small, the standards can be considered homogeneous.

One important pair of hypotheses is:

H_0: $\Delta \leq \delta$

vs.

H_1: $\Delta > \delta$

for some appropriate δ.

Other measures may also be considered, such as one based on the range.

Let: $R = \underset{k}{\text{Max}}\ \mu_k - \underset{k}{\text{Min}}\ \mu_k$.

Then another pair of hypotheses is:

H'_0: $R \leq \delta'$

vs.

H'_1: $R > \delta'$

It is not likely that any set of M standards can contain identical enrichments. Therefore, it may not be necessary to consider hypotheses that specify identical enrichments. We take δ and δ' to be strictly positive.

The choice of δ and δ' cannot be given by any a priori rule. The laboratory making the measurements will have to choose δ and δ'. If the standards do satisfy H_0 or H_0' they must be similar enough to fulfill their intended purpose, see Section 4.

SECTION 2. Some test statistics for the hypotheses.

Let PA(m,n) denote the peak area of the n-th replicate of the m-th standard, m = 1, ..., M, n = 1, ..., N. Let

$$\overline{PA} = \sum_{m,n} PA(m,n)/M \cdot N \text{ and } \overline{PA}(m) = \sum_{n} PA(m,n)/N \quad .$$

Then a good test (the uniformly most powerful invariant unbiased test if peak areas are assumed to be independent and normal) of H_0 vs. H_1 is given by using the noncentral F statistic

$$F(\Delta/\sigma^2,\ M-1,\ M(N-1)) = \frac{N\sum_{m} (\overline{PA}(m) - \overline{PA})^2}{\sum_{m,n} (PA(m,n) - \overline{PA}(m))^2}$$

This statistic has noncentrality parameter Δ/σ^2, and M-1 and M(N-1) degrees of freedom.

Any test of the form

Reject H_0 when $F > k$

do not Reject H_0 when $F \leq k$

is a good test. The constant k is chosen to provide a desired level of significance. In applications $M(N-1)\ \delta/\sum_{m,n}(PA(m,n)-\overline{PA})^2$ can be used as an estimate of the noncentrality parameter. The constant k can be chosen from the tables in [3].

A statistic for testing H'_0 vs. H'_1 is based on the range of the averages $\overline{PA}(m)$. Let M^* and M_* denote the indexes corresponding to the largest and smallest values of $\overline{PA}(m)$. If there are two or more indexes corresponding to the largest or smallest value an arbitrary choice may be made among them. Let the range of the statistics $\overline{PA}(m)$ be denoted by $\hat{R}$, i.e., $\hat{R} = \overline{PA}(M^*) - \overline{PA}(M_*)$. We take as an estimate for the variance of $\hat{R}$,

$$\hat{\sigma}^2(M^*,M_*) = \frac{2\sum_{m,n}(PA(m,n) - \overline{PA}(m))^2}{M(N-1)} .$$

By taking this choice we ignore the possibility that M^* and M_* do not correspond to the populations with the largest and smallest values of μ, respectively. No universally better choice is known by the author.

Then a statistic for testing H'_0 vs. H'_1 is:

$$T = \frac{(\hat{R}-\delta')}{\hat{\sigma}(M^*,M_*)}$$

In many applications there are hundreds of counts in each channel. Therefore the probability distribution of these counts can be well approximated by a normal distribution, see [4]. In these cases T has approximately the distribution of student's t with M(N-1) degrees of freedom, see also [5]. The test rejects H'_0 if

$$T > K'$$

and does not reject H'_0 if

$$T \leq K' .$$

The constant K' is obtained from any t table. It should be noticed that the indexes M^* and M_* may not correspond to the standards with the largest and smallest theoretical mean level μ. Therefore, the constant K' may have to be modified, see [6].

Notice that if either hypothesis H_0 or H'_0 is rejected then multiple comparison methods should be applied to find the "bad actors" among the group of standards, see [2]. An example illustrating the use of these tests is given in Section 4.

The next section deals with the interaction between the weights W and the test statistics.

SECTION 3. Finding extreme peak areas.

It is clear that the choice of weights, W, used in defining peak area affects the values of F and T. It would be wonderful if we could plot F and T as a function of W. However, j_2, the dimension of W, is usually much larger than 2, see Section 4. This implies that these plots are not practical.

Instead we settle for the comparison of a few choices of weights. The most important choices are those that the measurement laboratory believes produce meaningful and accurate peak areas, see [1]. In addition it may be useful to compute the set of weights which make the statistics F and T take on their largest and smallest values. Procedures for calculating these weights are given below.

The weights for the F statistic are provided first. Let Y_{jmn} denote the counts in the j-th channel of the n-th replicate of the m-th standard, $j \in J_1+J_2$, $1 \leq m \leq M$, $1 \leq n \leq N$. The usual $\cdot$ notation is used to denote an

average. The average over any set of subscripts is denoted by replacing the subscript(s) with a dot, e.g., $Y_{jm\cdot} = \sum_{n} Y_{jmn}/N$, $Y_{j\cdot\cdot} = \sum_{nm} Y_{jnm}/NM$.

Let the matrix A^t, where a superscript t denotes transpose, have its jm-th entry, a_{jm}, given by the equation $a_{jm} = Y_{jm\cdot} - Y_{j\cdot\cdot}$. This matrix has dimensions j_2 x M. In addition let the matrix B^t (j_2 x NM) have entries whose ((m-1)N+n)-th column is:

$$\begin{bmatrix} Y_{1mn} - Y_{1m\cdot} \\ \cdot \cdot \cdot \\ Y_{j_2mn} - Y_{j_2m\cdot} \end{bmatrix}$$

Recall that $W^t = (w_1, \ldots, w_{j2})$. The F statistic can now be rewritten as:

$$F = N \frac{W^tA^tAW}{W^tB^tBW}$$

The weights which maximize and minimize F subject to $\sum_{J_1} w_j = 1$ are characterized below. If the elements of B are normal and n > 2 then B^tB has full rank j_2. (The assumption of a normal distribution is only an approximation to the truth. There remains a very small chance that B^tB is not of full rank. If the rank of B^tB is not j_2, then the rest of the analysis in this section cannot be done.) It is known that B^tB can be rewritten as the square of a full rank symmetric matrix, Q, i.e., $B^tB = Q^2$, see [7].

Let $\mu = Q\ W$ (notice that μ does not denote any mean value). Then F satisfies the following equations:

$$F = NW^tQ(Q^{-1}A^tAQ^{-1})QW/\mu^t\mu$$

$$= N\mu^t(Q^{-1}A^tAQ^{-1})\mu/\mu^t\mu \ .$$

The constraint $\sum_{j\epsilon J_1} w_j = 1$ can be rewritten as:

$\mu^t\, Q^{-1}\, d = 1$ where d is a vector having ones in the first j_1 and zeroes in the last (j_2-j_1) places, i.e.,

$$d^t = \underbrace{(1 \ldots 1}_{J_1} \ \underbrace{0 \ldots 0)}_{J_2}$$

It is known that the μ vectors which maximize and minimize

$$\mu^t(Q^{-1}A^tAQ^{-1})\mu/\mu^t\mu$$

are the eigenvectors corresponding to the largest and smallest eigenvalues of $Q^{-1}A^tAQ^{-1}$ respectively, (see [7]).

Let $V^{(1)}, \ldots, V^{(j_2)}$ denote an orthonormal set of characteristic vectors (eigenvectors) for this matrix (see [7]).

The eigenvectors $V^{(i)}$, $V^{(j)}$ corresponding to largest and smallest eigenvalues satisfy

$$V^{(i)}Q^{-1}d \neq 0$$

$$V^{(j)}Q^{-1}d \neq 0$$

with probability one and are the solution, i.e., they maximize and minimize F. (We note that the independent eigenvectors need only be orthogonal, and need not be orthonormal. The F statistic is invariant to

scale changes.) The untransformed vectors $V^{(i)}$, $V^{(j)}$ are only of intermediate interest. The final choice of weights is obtained from

$$W = Q^{-1}V \ .$$

The study of these eigenvectors (weights) may be of independent interest. This is particularly true for the weights which minimize the F statistic. Our analysis is similar to principal components analysis, see [7]. In particular if identical standards are analyzed by the above analysis then the eigenvectors help to characterize the measurement process.

Next the weights for the T statistic are found. We use the matrix Q defined in the last derivation. Consider the M(M-1) vectors $D_{mm'} = Q^{-1}\begin{bmatrix} Y_{1m\cdot} - Y_{1m'\cdot} \\ \cdot\ \cdot\ \cdot \\ Y_{j_2m\cdot} - Y_{j_2m'\cdot} \end{bmatrix}$ for all m and m′. Let $\mu = Q\,W$ as before. Notice that T is proportional to $\max \dfrac{D^t_{mm'}\,\mu}{(\mu^t\mu)}$. Then the vector μ which maximizes T is the vector $D_{mm'}$ of greatest length. (Recall that the constraint $\sum_{1}^{J} W = 1$ serves only to rule out the null vector.)

The weights that minimize T satisfy the following quadratic programming problem:

$$\underset{W}{\text{Min}}\ Z$$

such that for all m and m′

$$D^t_{mm'}\,\mu \leq Z \quad \text{and}$$

$$\mu^t\mu = 1 \quad \text{where}$$

$$\mu = QW.$$

Discussion.

The largest and smallest values of a set of F and T statistics do not have F and t distributions. The distributions for the F statistic case are given in [7] under H_0 with $\delta = 0$. The distributions for the T statistic case have yet to be worked out, and according to M.A.H. Dempster [8] distributions for stochastic programming problems are complicated.

SECTION 4. An F and T Statistic Example.

The National Bureau of Standards (NBS) is participating in a certification program for U_3O_8 low enrichment (<3%) standards. At the time of this writing NBS has only analyzed prototype standards. There are three replicate measurements for each of twelve prototype standards. The three measurements for one peak and its backround region, for one of the samples, are shown in Figure 2. The peak area was computed using the sum rule, i.e., $W_1 = W_2 = \ldots = W_{j_1} = .1$ and $W_{j_1+1} = W_{j_1+2} = \ldots = W_{j_2} = -.1$, where $j_2 = 2j_1$. The value of the F statistic for this peak area formula is shown in Table 1 along with the individual means PA(m) for each of the prototype standards. The value of the T statistic is 63.7. These statistics correspond to real differences among the standards of about (10^{-3})% enrichment. On the basis of both statistics we can say the prototype standards are different but good enough for their intended use.

The weights corresponding to the extreme values of the F and T statistics were not computed. However it is conjectured that the weights corresponding to the maximum of these statistics will give large weight to anomalies in the data. It is also conjectured that if the prototype standards are nearly equal

then the weights which minimize the statistics should give appropriate peak area rules.

SECTION 5. Additional Considerations.

In addition to the T and F other statistics may be worth computing. These include rank tests, see [9]. For testing H_0' vs. H_1' the maximum likelihood ratio statistic is known to be good when N is large.

The minimizing and maximizing weights for the T statistic may not be the only weights of interest. For exploratory purposes it will be useful to choose weights orthogonal to the previously chosen ones, which maximize the T statistic. Finally it may be worthwhile to track the optimum weights over time, since they are directly related to the necessary background correction.

SECTION 6. Conclusion and Summary.

Two important statistics, F and T, were given for testing the sets of hypotheses (H_0 vs. H_1) and (H_0' vs. H_1'). These hypotheses are related directly to two possible measures of the homogeneity of the standards. For these statistics procedures were given for finding weights which make the standards look most and least homogeneous. The F and T statistics from these weights can be used as a basis for conservative decisions about the standards. For example, the minimum value found for the T statistic is at least as small as the value of T that would be found using a "proper" choice of peak area weights. Thus if the minimum value of T is large, one can be sure that no linear peak area formula could reduce an apparent inhomogeneity of standards.

Acknowledgement: I thank J. R. Rosenblatt for many helpful suggestions. The data in Section 4 were jointly analyzed with D. Hogben.

References

[1] Helmer, R. G., Lee, M. A. (1980). Analytical Functions for Fitting Pears from GE Semiconductor Detectors, Nuclear Instruments and Methods 178, 499-512.

[2] Scheffe´, H. (1959). The Analysis of Variance, John Wiley, New York.

[3] Tiku, M.L. (1967). Tables of the Power of the F-Test, JASA 62, 525-39.

[4] Feller, W. (1968). An Introduction to Probability Theory and Its Applications-3rd Edition, Wiley, New York.

[5] Beyer, W. H. (1968). Handbook of Tables for Probability and Statistics, Second Edition, Part VIII, The Chemical Rubber Co. (CRC), Cleveland.

[6] Gibbons, J.D., Olkin, I., and Sobel, M. (1977). Selecting and Ordering Populations, John Wiley, New York.

[7] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis, John Wiley, New York.

[8] Dempster, M.A.H. Ed. (1974). International Conference on Stochastic Programming, I. Introduction, Oxford, New York.

[9] Hajek, J. and Sidak, Z. (1967). Theory of Rank Tests, Academic Press, New York.

table 1

I ANALYSIS OF VARIANCE I

SOURCE	D.F.	SUMS OF SQUARES	MEAN SQUARES	F RATIO	PROB.
BETWEEN GROUPS	11	1862338600	169303510	358.531	.000
SLOPE	1	136082740	136082740	2.663	.116
ABOUT LINE	10	1726255900	172625590	365.566	.000
WITHIN GROUPS	24	11333147.	472214.44		
TOTAL	35	1873671700			

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCE BETWEEN MEANS
SIGNIFICANCE LEVEL IS APPROXIMATELY .000

I ESTIMATES I

GROUP	NO.	MEAN		MINIMUM		MAXIMUM		SUM RANKS
1	3	313653.33		312766.00		314327.00		87.0
2	3	296557.00		296031.00		297184.00		16.0
3	3	305500.33		304649.00		305937.00		50.0
4	3	317793.00	H	317216.00	H	318494.00	H	105.0H
5	3	309431.66		308581.00		310178.00		63.0
6	3	310640.00		309866.00		311380.00		75.0
7	3	309793.66		308770.00		310405.00		69.0
8	3	295899.33	L	294959.00	L	296879.00	L	9.0L
9	3	316038.00		315500.00		316703.00		96.0
10	3	304434.33		304028.00		304675.00		43.0
11	3	296897.33		296299.00		297307.00		20.0
12	3	303975.00		303935.00		304012.00		33.0
TOTAL	36	306717.75		294959.00		318494.00		

GROUP	WITHIN S.D.		S.D. OF MEAN		95 PCT CONF INT FOR MEAN
1	802.13486		463.11278		311660.70 TO 315645.96
2	583.09776		336.65165		295108.49 TO 298005.51
3	737.35835		425.71404		303668.61 TO 307332.05
4	647.96065		374.10025		316183.36 TO 319402.64
5	803.59587		463.95629		307435.40 TO 311427.93
6	757.57244		437.38465		308758.07 TO 312521.93
7	892.08762		515.04703		307577.57 TO 312009.75
8	960.60415	H	554.60506	H	293513.04 TO 298285.63
9	611.47281		353.03399		314519.00 TO 317557.00
10	353.88746		204.31702		303555.22 TO 305313.45
11	529.82292		305.89341		295581.17 TO 298213.50
12	38.587563	L	22.278539	L	303879.14 TO 304070.86
MODEL					
FIXED	687.17861		114.52977		306481.38 TO 306954.13
RANDOM	7512.2900		2168.6113		301944.66 TO 311490.82
UNGROUPED	7316.6575		1219.4429		304242.15 TO 309193.35

Figure 1

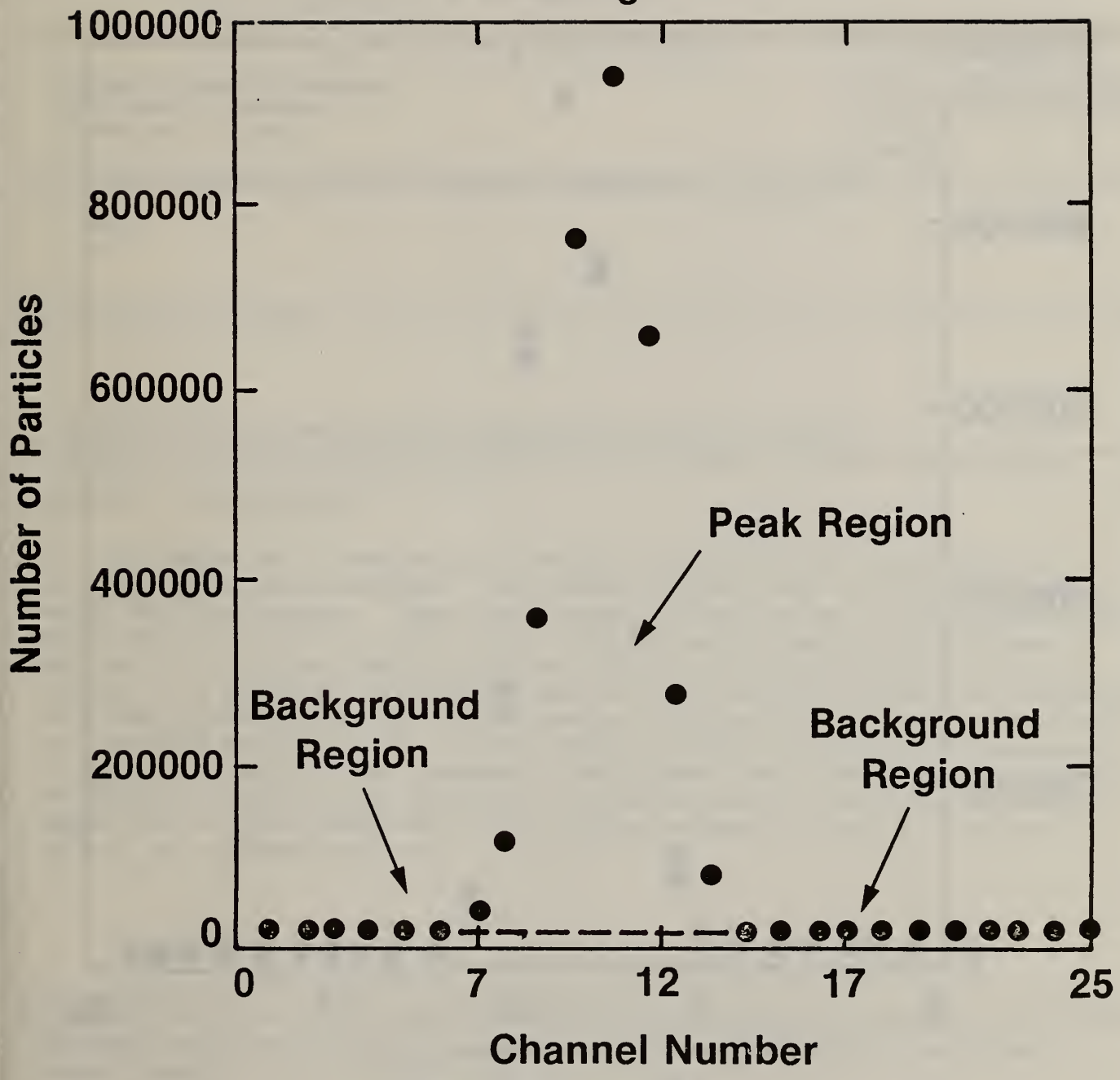
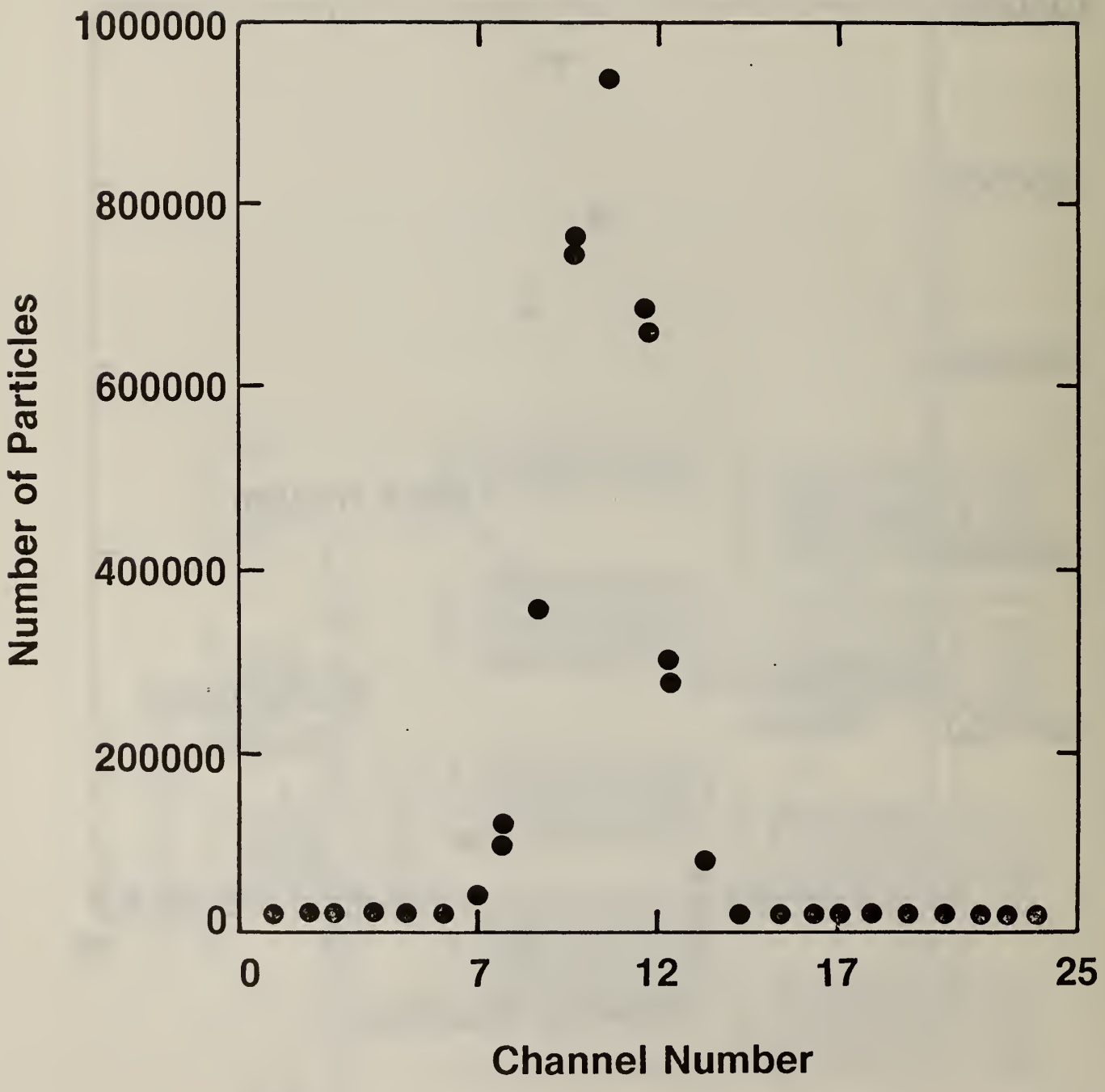

1000000
800000
600000
400000
200000
0
Number of Particles
Peak Region
Background Region
Background Region
0
7
12
17
25
Channel Number

Figure 2


1000000
800000
600000
400000
200000
0
Number of Particles
0
7
12
17
25
Channel Number

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBSIR 81-2222	2. Performing Organ. Report No. 714	3. Publication Date

4. TITLE AND SUBTITLE

Title: Some Effective Statistical Approaches for Presenting Interlaboratory N.D.A. Enrichment Measurements

Subtitle: Presenting N.D.A. Enrichment Measurements

5. AUTHOR(S)

Clifford Spiegelman

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

NRC AT(49-25)-9009

8. Type of Report & Period Covered

NBSIR

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

NRC

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

SECTION 1. Introduction.

It is important to realistically evaluate interlaboratory measurements on prototype and primary standards. The emphasis in this paper is on nondestructive assay (NDA) measurements. These measurements are based on an estimate of the total number of counts from a target source. This estimate is called the peak area calculation. It is necessary to distinguish real differences between laboratory findings from artificial differences such as those caused by artifacts of individual peak area calculations. This report provides guidance for assessing the magnitude of these peak area artifacts. A wide variety of statistical tests for homogeneity of the peak areas for these standards is considered. For these tests, algorithms are provided for finding linear peak area computations which either make the standards appear most or least homogeneous. These tests include the usual F statistics as well as the standardized range.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

Principal components; multiple comparisons; multivariate analysis, peak areas, minmax, maxmin

13. AVAILABILITY

☒ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

18

15. Price

$5.00